STUF, Akademie Verlag, **65** (2012) 1, 78–94

Jean Léo Léonard (Paris), Vittorio dell'Aquila (Vasa)
& Antonella Gaillard-Corvaglia (Paris)

The ALMaz (*Atlas Lingüístico Mazateco*): from geolinguistic data processing to typological traits

Abstract

Mazatec (Popolocan, Eastern Otomanguean) became world-renowned following an article by Pike & Pike (1947) and the famous chapter VIII in Pike (1948) on the Huautla variety, which inspired several other seminal studies by Bull (1984), Kirk (1966), Jamieson (1988, 1996), and more recently by Steriade, Golston & Kehrein (1998, 2004), Silverman et al. (1995) on this extremely relevant language as far as phonological typology is concerned. However the early monographs and sketches, which have had a major impact on modern linguistics (laying the premises for the syllabic constituency theory, the theory of tones and tone sandhi as well as their functions in inflectional systems), only take into account a minute proportion of this language's inner diversity. The ALMaz project is an attempt to both revisit second-hand data on Mazatec varieties all over the area where it is densely spoken (especially using Kirk 1966 lists of over 700 cognates as a data base), and to gather and process new data on Mazatec diatopic and diastratic variation, using computational geolinguistics.

Keywords: Mazatec, phonology, typology, breathy, creaky, computational, geolinguistics

1. Mazatec's relevance

From the standpoint of the historiography of modern linguistics, Mazatec belongs to a very specific group among the world's languages: those which have made breakthroughs possible in various fields of research – the *paradigmatic languages*, in Kuhnian terms.

Firstly, Eunice & Kenneth Pike's 1947 paper on the "immediate constituents of Mazatec syllables" (Pike & Pike 1947), published in the *International Journal of American Linguistics*, is alleged to be a milestone for theories of Syllable Constituency, which assume that any syllable is made of an onset and a rhyme (a nucleus and an optional coda). Secondly, Mazatec turned out to be one of the main token languages, along with Mixtec (*tu'un savi*), for the study of tone patterns in grammar in Pike's seminal essay on tone languages (Pike 1948). A whole chapter is dedicated to the function of tone in the verbal system of Huautla Mazatec (Pike 1948: 95–161), providing a complete survey of tone in its link with grammatical structure, rather than taking tone into account as a property of the lexicon. Thirdly, Sarah Gudschinsky's paper on Mazatec dialect history (Gudschinsky 1958b) should be familiar to dialectologists as one of the most perfect "studies in miniature", as the author herself put it, of dialectical variation in space and time for an Amerindian language. Not only did Gudschinsky put forth a classification for Mazatec dialects based on an elaborate phonological taxonomy, dealing with the somewhat uncanny sound changes such as $t^{j}k > hk$, $tk > sk$, $tk > hk$), but she suggested an elegant chronology of settlements, inter-ethnic contact, and the diffusion of town dialects over historical areas. This

stands as a remarkable achievement among geolinguistic descriptions of a language's inner diversity.

Fourthly, Pike & Pike's (1947) theory of complex onsets in Mazatec was still the issue of a crucial debate in phonological theory fifty years later, on the typological relevance of features such as *breathy* and *creaky* in the world's languages (Ladefoged & Maddieson 1996: 314–320, Kirk 1993, Silverman et al. 1995, Golston & Kehrein 1998). Pike & Pike's extensive sequences of consonants in Mazatec complex onsets were revisited according to cross-linguistic and dialectical evidence from other Mazatec varieties (particularly from the innovative dialect of San Felipe Jalapa de Diaz). As opposed to the tenants of the complex constituent theory, typologists have pointed out the fact that breathiness and creakiness distributed between onsets and nuclei may provide a better account for surfacing segmental complexity in Mazatec than the cluster analysis (e.g. *ht?, ?nt, hnt, nth, ntsh, nkh* as *t̤V̰ n̰t*, *n̤t*, *nt̤V̰ nts̤V̰ nk̤V̰*).

A minimalist description of Mazatec phonological system would go as follows: no oral labial stops (except [b] as consonantized /w/). The consonant inventory can be summed up in a few words: stops *t, k, ?, nt, nk, f* (*<= w + h, h + w*); fricatives *s, ʃ, h*; sonorants *m, n, w, l* (*<nt*^{j}). Glottal constriction? and friction *h* combine to the right or to the left of any of these basic consonants. The vowel system is roughly pentavocalic (*i, u, e, o, a*), with restrictions on *u* and/or *o* according to the dialect (Chiquihuitlán has both, whereas Huautla only has *o* as a lowered *u* merger). Two main parameters make the inventory complex: a *voice quality correlation* (modal vs. breathy and creaky vowels) and a nasal/oral correlation – both can accumulate. The status, value and function of diphthongs, as we shall see, is still a topic of debates. Last but not least, as in most Meso-American languages, Mazatec has a rich (and fairly unstable) system of complex nuclei known as "rearticulated" or "interrupted vowels" (homorganic chains of the VhV and V'V type), which tend to delete their first peak in the Midland dialects (especially in Jalapa). The interplay of *h* and *?* in consonant clusters, in morphonological hyphenisation (as to *?*), or as the product of preconsonantal spirantisation of stops, of vowel breathiness, of rearticulated vowel syncopation, etc. conspire to make Mazatec postlexical realizations look quite intricate, whereas basic sound patterns and categories may actually be quite simple, as suggested with the minimal consonant inventory above. A CV canonic shape can be postulated for the Mazatec syllable, without any further complication, in spite of the surface complexity of onsets described in Pike & Pike (1947).

Mazatec still badly needs scrutiny and documentation. Some dialects have been thoroughly surveyed in monographs, such as Chiquihuitlán (Jamieson 1982, 1988, 1996). There are some extremely valuable fragments of phonology, grammar, or the lexicon for the dialects of Huautla (Cowan 1965), San Jerónimo Tecoatl (Bull 1984), Jalapa[1] and San Miguel Soyaltepec (Pike 1956). Gudschinsky's comparative sketch for Proto-Popolocan (1958a) and Paul Livingston Kirk's Mazatec comparative phonology (1966) provide an overview of the structural diversity of akin languages (ixcatec, chocho or ngigua, Popoloca) and of Mazatec dialectical patterns. Nevertheless, monographs on inflectional morphology, syntax and the lexicon are still missing for such important dialects as San Miguel Soyaltepec, San Pedro Ixcatlán, San Felipe Jalapa de Días, San Lorenzo Cuaunecuiltitla. The data currently available should be first processed and revisited within the scope of any large

[1] See Moreno (2008) for a collection of bilingual texts and an outstanding bilingual glossary of Jalapa Mazatec.

scale project for the documentation of Mazatec dialects. This paper aims at providing a glimpse of what can be achieved by going over formerly published data, using current technologies for data mapping from the standpoint of geolinguistics.[2]

2. The original data

Our atlas is entirely based on KIRK's (1966) comparative monograph. We shall first explain where we started, working out our own atlas of Mazatec dialects according to KIRK's data. A further step, in progress, consists of going back over the whole process of collecting data, through fieldwork in the same spots surveyed nearly fifty years earlier by PAUL KIRK for his Ph.D. dissertation, with a much wider question-list, taking into account over 2,500 items in phonology, morphology, syntax and the lexicon.

KIRK's original work includes 6633 locally attested words collected in 12 villages and grouped into 731 etyma. A hand drawn map of the traditional Mazatec area highlights the river network and shows the main routes available at that time, with the location of the towns or *municipios* visited by PAUL KIRK (see Figures 1.1. and 1.2).

The linguistic data, i.e. the 6633 single words, is grouped by etymon in several hundred tables to be found throughout KIRK's Ph. D. dissertation (KIRK 1966). All the etyma and the author's statements on historical phonetics of the Mazatec dialects are justified and exemplified by KIRK through the tables and in detailed discussion.

In order to create an atlas out of the dialect sample given by KIRK in his Ph. D. dissertation, we first digitalised the tables by retyping them in spreadsheet. As an example, the table in Figure 2 was digitalised as shown in Figure 2.2.

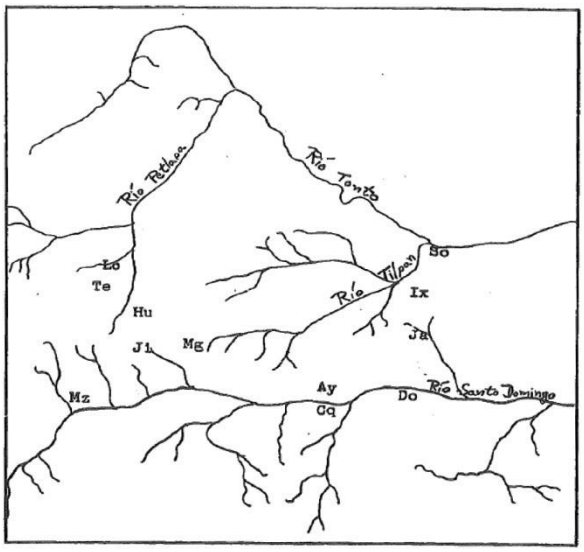

Figure 1.1: Towns or *municipios* visited by PAUL KIRK for his Ph. D. dissertation (KIRK 1966)

[2] We dealt more extensively with Mazatec phonology, written and oral elicitation of Mazatec dialectical data (cross-elicitation), applied dialectology and the use of available documentation on Mazatec in LÉONARD (2010).

CHART 2

SYMBOLS USED FOR THE MAZATEC LANGUAGES

(Gudschinsky 1953, 1956, 1959; Kirk 1966)

	1966	1959	1956	1953
Mazatlán de Flores	Mz	Mz	Maz	
San Bartolomé Ayautla	Ay			
San Juan Chiquihuitlán	Cq	Ch	Ch	
Jalapa de Díaz	Ja			
Santo Domingo del Río	Do			
Huautla de Jiménez	Hu	H	H	H
Santa María Jiotes	Ji		MJ	
San Miguel Soyoltepec	So	S	S	S
San Pedro Ixcatlán	Ix		I	I
San Miguel Huautla	Mg		M	SM
San Lorenzo Cuaunecuiltitla	Lo			
San Jerónimo Tecoatl	Te	SJ	T	T

Figure 1.2: Kirk's localities and previous surveys (Kirk 1966: 3)

(1) Preceding *u, *ʔñ is reconstructed from Mz ʔñ :
Ay ʔñ : Cq ʔñ : Ja ʔɨ : Do ʔñ : Hu ʔñ : Ji ʔñ :
So ʔñ : Ix ʔñ : Mg ñʔ : Lo ʔñ : Te ñ as in set 339
*nɨ⁴ʔñɨ́⁴, *nạ⁴ʔñɨ́⁴ teeth:

Mz	nɨʔñɨ́	Ji	nɨʔñɨ́
Ay	nteʔñá	So	nạ⁴ʔñɨ́⁴
Cq	nẹɨʔñɨ́	Ix	nɨʔñɨ́
Ja	nɨ²ʔɨ́²	Mg	nẹñʔɨ́
Do	nɨʔñɨ́	Te	nɨñɨ́
Hu	nɨ⁴ʔñɨ́⁴, nɨ³ʔñɨ́⁴		

Additional examples: 200, 252, 715, 721, and 722.

Figure 2.1: A sample of Kirk's tables: data and comments for cognate set [1] (Kirk 1966)

pMaz	English	Mz	Ay	Cq	Ja	Do	Hu	Ji
*nĩ′ñũ	'teeth'	nĩ′ñũ	nte′ñã	nẹnĩ′ñũ	nĩ′ñũ	nĩ′ñũ	nĩ′ñũ	nĩ′ñũ

So	Ix	Mg	Te
nã′ñũ	nĩ′ñũ	nẹñ′ũ	nĩñũ

Figure 2.2: Kirk's cognate set [1] in tabular form for data processing, 2011

The original geographical data was compared with the demographic data provided by INEGI. According to the census data (year 2000) the Mazatec-speaking area includes 25 municipalities (municipios) for a total population of about 250,000 inhabitants of which about 200,000 declare being able to speak an indigenous language. KIRK managed to visit 12 towns in spite of the precarious conditions for transportation which prevailed at that time in the area. From our own fieldwork in August 2010 and ANTONIA COLAZO-SIMON's survey of seven dialects in January and February 2011, it seems that KIRK could only have had but a glimpse of the dialectal diversity of Mazatec. In the years to come, the ALMaz (*Atlas Lingüístico Mazateco*) network will be extended in order to include varieties from the 25 *municipios*, within the framework of our project on Otomanguean and Mayan geolinguistics (the MAmP project, i.e. *Meso-American MorphoPhonology* 2009–2013, supported by the Institut Universitaire de France)[3]. In each locality, speakers of both genders and representing both rural and urban areas should be interviewed.[4]

Municipality	Population	% indigenous speakers
Huautepec	6 535	98,9
San Pedro Ocopetatillo	877	98,7
Santa María la Asunción	3 339	98,7
San José Tenango	19 957	98,3
Santa Ana Ateixtlahuaca	524	98,0
San Lucas Zoquiápam	7 207	97,3
San José Independencia	4 538	97,2
San Bartolomé Ayautla	3 824	97,1
Santa María Chilchotla	21 398	96,6
San Lorenzo Cuaunecuiltitla	737	96,3
Santa Cruz Acatepec	1 266	95,8
Eloxochitlán de Flores Magón	4 136	95,3
San Pedro Ixcatlán	10 848	94,2
San Felipe Jalapa de Díaz	23 191	93,7
San Francisco Huehuetlán	1 379	93,4
Huautla de Jiménez	31 103	93,4
Mazatlán Villa de Flores	13 933	93,1
San Mateo Yoloxochitlán	2 925	88,8
San Jerónimo Tecoátl	1 701	88,8
San Miguel Soyaltepec	35 853	74,5
Chiquihuitlán de Benito Juárez	2 497	70,5
San Juan de los Cués	2 489	41,0
San Pedro Teutila	4 162	24,3
Santa María Tecomavaca	1 812	18,3
Acatlán de Pérez Figueroa	44 509	13,2

Table 1: Demographic data on the Mazatec area (Mexican census, 2002)

Current geodesic tools such as Google Earth enable more accurate maps of the Mazatec area, taking into account administrative divisions as much as orographical or hydrographical patterns.

[3] The MamP project also includes research on the historiography of 20th century research on inflectional morphology of Otomanguean languages (see LÉONARD & KIHM 2010).

[4] Although PAUL KIRK's data look fairly good and are indeed reliable, hardly any detailed information on his speech consultants can be found in his dissertation.

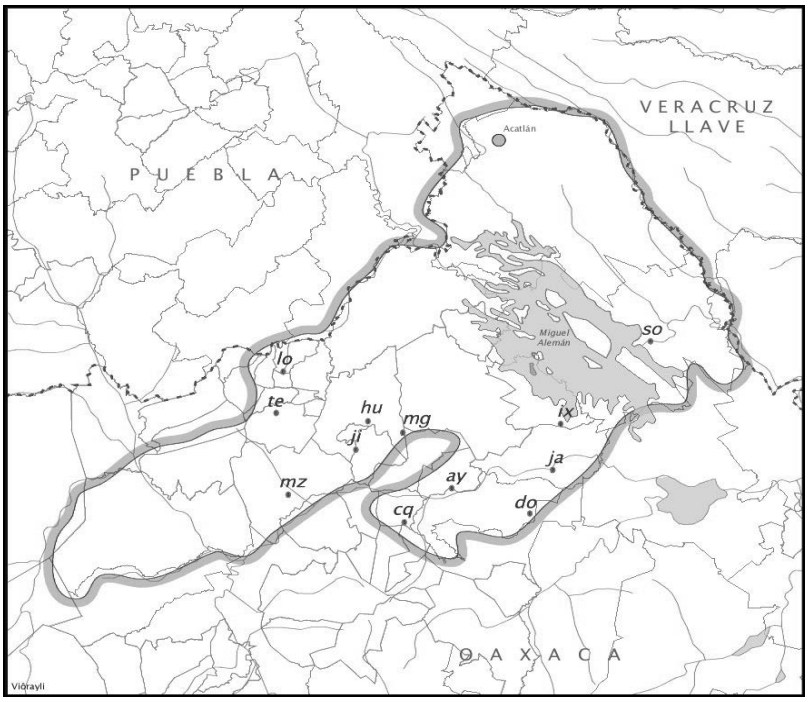

Figure 3: Kirk's survey of Mazatec dialects: *municipios* and administrative divisions

3. The database

In order to search through the dialectal data, to sort the results and to draw up linguistic maps from these, a database was developed. The database was fed by importing the previously prepared spreadsheet into various tables.

The database, displaying materials according to Kirk's phonetic transcription or its conversion according to the I.P.A. standards (see Figure 4), allows any user not only to submit specific and punctual queries, but also to create complete lists of words, generating as many small dictionaries or *microdictionaries* as one could wish for (see Figures 4.1–4.4 below), according to structural criteria. These lists are structured by default and alphabetically sorted by etymon (the Proto-Mazatecan roots as defined by Kirk), by English translation (original), Spanish translation (our integration) and by single item (tagged according to location).

Figure 4.1 shows a sample of a *microdictionary* generated by the database: an etymological dictionary of Mazatec, according to Kirk's data 1966.

Figure 4.2 shows a bilingual microdictionary Spanish-Mazatec, which could prove to be very useful to schoolteachers involved in Intercultural and Bilingual Education (EBI)[5]. There are many ways that microdictionaries, as by-products of the linguistic atlas proper,

[5] The ALMaz is a cooperative project developed for the benefit and with the contribution of Mazatec-bilingual teachers, with a strong concern for *applied dialectology* (see Léonard 2010).

Mazatec

Usa

Formato: pMaz

Record: 588

Totale: 730

Non ordinati

Id	pMaz_root1	pMaz_root1_ipa	pMaz_root2	pMaz_root2_ipa	pMaz_eng	pMaz_esp
552	sunnau	ʃuɲɲau			lime	limon verde
553	šu	ʃu			landslide	derrumbe
554	su	su			boils	hierve
555	šuhmā	ʃuhmā			seed	semilla
556	šuñū	ʃuɲū	šañū	ʃaɲū	dew	rocio
557	šuwi	ʃuwi			termite	termita
558	šu'u	ʃuʔu			navel	ombligo
559	šu'wa	ʃuʔwa	ša'wa	ʃaʔwa	lot	solar
560	šuhū	ʃuhū			paper	papel, carta
561	šuta	ʃuta	čuta	ʧuta	person	persona
562	šuti	ʃuti	čhuti	ʧhuti	tomato	tomate
563	šu'we	ʃuʔwe	ša'we	ʃaʔwe	wasp	avispa
564	taha	taha	ntaha	ntaha	tough, hard	duro
565	ta-	ta-			no longer	no más, temprano
566	taku	taku	tuku	tuku	show me	¡enséñame!
567	tāū	tāū			money	dinero
568	tencu	tentsu			goat	chivo
569	te	te			dances	baila
570	te	te			ten	diez
571	te	te			brilliant	brillante

Figure 4: A glimpse at the database, with KIRK's original notation and conversion according to I.P.A. conventions

could be contrived for the needs of scholars as much as for the benefit of local schoolteachers, as shown in Figures 4.3 and 4.4: an English-Mazatec version of the database and a dictionary according to *municipios*. Data of any of the twelve municipios could also be extracted in order to provide local Spanish-Mazatec or Mazatec-Spanish glossaries of Jalapa, San Lorenzo or Ayautla dialects.

4. A geolinguistic Atlas of Mazatec

4.1. Generating maps

The ALMaz digitalised dialectal corpus works like an automatic atlas, providing the user with a great number of geolinguistic maps. This interactive dimension is one of the main characteristics of the ALMaz. The descriptive maps – i.e. the maps on which raw original data can be localised on a base map – are created automatically out of the data already stored for microdictionaries. Any lemma and the associated forms in KIRK's cognate sets can be visualised on seven *base maps*: a general one in full colours displaying political and physical elements, a second one in which the physical characteristics of the territory are displayed in black and white in order to better pinpoint the linguistic information, two simple ones in black and white for printing, a political one with municipal borders, and two demolinguistic base maps in which the linguistic data are superposed onto choropleth maps showing the traditional distribution of the languages in contact in the Mid Papaloapán Bassin where Mazatec is spoken (namely Nahuatl, Mixtec, Cuicatec, Chinantec and Spanish).

Mazateco

***tsa?u** *v. 3s.* limps; cojea ◇ ay **khits?o–**, ja **tihwikits?ō**, do **tihwikits?o**, te **hekitsiu**, ix **ts?ihwikitsu?u** *v. 1s.* limp; cojeo, ix **ts?ihwikitsu**

***tsakhā?a**$^{1-13}$ *v. 3s.* stutters; tartamudea ◇ mz **sakha?a**, ay **tsakha?a**, ja **tsàkhā?a** $^{M-HM.}$ hu **tsakhā?a** $^{L-HL.}$ so **tsakhā?a**, ix **tsakhā?a**, te **tsakhā?a**

***tsanka**$^{3-42}$**/*thukha**$^{3-42}$; ◇ ay **timānkantsanka**, cq **tiwānka**, ja **tiwānkàntsànkà**, do **tiwānkatsanka**, so **thokhā** $^{L-LH.}$ ix **ts?amānka**, lo **tsonko**, te **tiwanka**, hu **mānka** *v. 3s.* cheses flees; mata piojos, hu **tuk n thuk** $^{L-BL}$ *v. 3s.* runs; curre

***tsankhū**$^{3-1}$**/*tsunkhū**$^{3-1}$ *v. 3s.* fears; tiene miedo ◇ mz **tsakhū**, ay **tsakhū**, cq **tsakhū**, ja **tsànkhū** $^{M-H.}$ do **tsankhū**, hu **tsukhū** $^{L-H.}$ ji **tsukhu**, ix **tsakhū**, mg **titsakhū**, lo **tsakhī**, te **titsakhū**, so **tsakh re** $^{L-L-MB<L-MH-B.}$ so **wetjk** $^{LH-LM}$

***tsa?u** *v. 3s.* goes out (fire); sale (fuego) ◇ hu **whits?au** he puts out; pone afuera, ji **kawits?o** he puts out; pone afuera, so **kits?o** he puts out; pone afuera, ix **kits?uja**, mg **hwets?o** he puts out; pone afuera, lo **kitsu?uu**, te **kitsiu ni?ī**

***tsa**^{32} *conj.* if; si ◇ ay **tsa**, cq **tsa**, ja **tsà**, do **tsa**, hu **tsà** $^{L\ EXPECTED.}$ so **tsà** $^{LM.}$ ix **tsa**, lo **tso**, te **tsa**

***tsa** *adj.* fast; rápido ◇ ay **kitsa**, ja **kitsa**, do **kitsa**, ji **kitsa**, ix **tsa**, lo **kimōtsa–**

***tsakhī**$^{4-1}$ *s.* girl; chica, muchacha ◇ ja **tsakhī** $^{L-H.}$ do **tsakhī**, so **tsakhī**, ix **tsakhī**

***tsanka** *s.* cotton; algodón ◇ mz **tsanka**, ay **tsanka**, cq **tsanka**, ja **tsanka**, do **tsanka**, hu **tsanka**, ji **tsanka**, so **tsanka**, ix **tsanka**, lo **tsonko**, te **tsanka**

***tse?e**$^{1-1}$ *s.* lazy; haragán ◇ ay **tse?e**, cq **tse?e**, ja **tse?e**, do **tse?e**, hu **ts?ē** $^{H.}$ so **tse?e**, ix **tsi?e**, lo **ʧotsa?a**, te **tse**, mz **ts?i**, mz **ts?e**

***tse** *adj.* big; grande, grande ◇ hu **tse**, ji **tse**, so **tse**, ix **tse**, mg **tse**, lo **tsa**

***tse** *adj.* full; lleno ◇ mz **kitsi**, ay **kitse**, cq **kitse**, ja **kitsè**, do **kitse**, hu **kitse**, so **kitse**, ix **kitse**, mg **kitse**, lo **kitsa**, te **kitse**

***tse** *v. 3s.* lights (a fire); enciende (lumbre) ◇ ay **tetse**, cq **tsitse**, ja **tètsè**, do **kwetse**, hu **w?ētse**, ix **watse**, te **tetsi**

***–tse** *adj.* new; nuevo ◇ mz **ʧatse**, ay **tsatse**, cq **ʃatsa**, ja **tsetse**, do **tsetse**, hu **ʧutse**, ji **thatse**, so **thatse**, ix **thatse**, mg **ʧatse**, lo **ʧotsa**, te **ʃitse**

***tsehē**$^{4-2}$ *s.* coyote; coyote ◇ mz **tsihē**, ay **thiutsehē**, hu **tsē** $^{BM.}$ mg **ntsehē**, lo **tsahā**, te **tsihē**

***tse?e** *pron.* his; su ◇ mz **ts?e**, ay **kwitse?e**, ja **kwits?e** $^{H-L.}$ do **kwits?e**, hu **ts?e**, ji **tse?e**, so **tse?e**, ix **tsi?e**, lo **tsa?a**, te **tse**

***tsha**^{2} *s.* coati; coati, cusumbe, cuchucho, guache ◇ ay **thiutsha**, cq **ʧutsha**, ja **tshá**, do **tsha**, so **tshā** $^{MH.}$ ix **tsha**, lo **tsho**, te **tsha**

1

Figure 4.1: A sample of an etymological *micro-dictionary*, according to Kirk's data (1966)

Mazateco

able *adj.* ay **ʧukwhare**, ja **ʧukwha** $^{H-L.}$ do **ʧukwha**, hu **hʧukhua** $^{H-B.}$ ix **ʧukhua**, mg **ʧukwha**, lo **ʧikwha**, te **ʧikhua**.

ables *v. 3s.* mz **akwimā–**, ay **mā–**, cq **mā–**, ja **mā** $^{ML.}$ do **mā–**, hu **mā–**, so **kumā–** $^{L-H.}$ ix **mā**, lo **mō**, te **mā–**.

above *adv.* mz **khisū**, ay **khihŋāsūja**, cq **nkasū**, ja **khihŋāsū** $^{M-M-M.}$ do **khihŋāsū**, hu **nkasū** $^{L-M.}$ ji **sū?nka**, so **nkasū** $^{B-H.}$ ix **khihŋāsū**, lo **tihŋōʃw**, te **isūnka**.

abstract thing *s.* hu **khua** $^{B.}$ ji **khuatho**.

accustomed to *adj.* mz **kamānka**, mz **kamānka**, ay **kamānka**, cq **kamānka–**, ja **timānka** $^{H-M-ML.}$ do **kamānka**, hu **mānkali** $^{L-L-B.}$ ix **kamānka**, mg **mānka**, lo **jakhonka**.

acid *adj.* mz **sā**, ay **sā**, cq **sā**, ja **sā** $^{M.}$ do **sā**, hu **sā**, ji **sā**, so **sā**, ix **sā**, mg **sā**, lo **sō**, te **sā**.

across *adv.* ay **nkuwaha**, cq **kuankeikhā**, ja **ngawakha** $^{M-M-H.}$ ix **nkuwakha**, lo **ko?okho**.

adobe *s.* mz **ʃinti**, ay **ʃi?nte**, cq **ʃi?nte**, do **ʃhinte**, hu **ʃi?nte** $^{B-B.}$ ji **ʃi?nte**, so **nīntethāī**, ix **ʧhinte**, mg **ʃinte**, lo **ʃi?nta**, te **ʃinte** mz **ʃinti**, ay **ʃi?nte**, cq **ʃi?nte**, do **ʧhinte**, hu **ʃi?nte** $^{B-B.}$ ji **ʃi?nte**, so **nīntethāī**, ix **ʧhinte**, mg **ʃinte**, lo **ʃi?nta**, te **ʃinte**.

afternoon *s.* mz **nkiʃura**, ay **kwankiʃō**, cq **nkiʧū–**, ja **giʃū** $^{M-H.}$ do **giʃū**, hu **māʃū** $^{L-M.}$ ji **nkiʃū**, so **nkisū** $^{L-H.}$ ix **kamānkiʃū**, lo **kiʃī**.

afterwards *adv.* mz **haskā**, ay **haskā**, cq **nkaskā**, ja **iskā** $^{L-M.}$ do **haskā**, hu **haskā** $^{B-L.}$ ji **askānī**, so **haskā** $^{B-L.}$ ix **haskā**, lo **kosko**, te **iskā**, te **kiskā**.

alive *adj.* ay **tihnākū**, cq **kahwekakū**, ja **tikū** $^{H-M-}$ do **tikū**, hu **kū**, ji **kū**, ix **tikū**, lo **tinōkī**, te **k** , te **tihnāk** .

alligator *s.* ay **ʃkha**, ja **ʃkha** $^{M.}$ do **kha**, hu **ʃka**, so **ʃka**, ix **ʃkha**, mg **ʃka**, lo **ʃko**, te **ʃka**.

alone *adj.* mz **ʃutakā**, mz **sawa**, ay **k?a**, ay **suwa**, cq **kāt̄ā**, cq **kamāsuwa**, ja **k?ā** $^{M.}$ ja **suwa** $^{M-M.}$ do **suwa**, do **k?ā**, do **kāt̄ā**, hu **k?ā**, hu **sua** $^{BM.}$ so **sawa** $^{L-H.}$ ix **k?ā**, ix **suwa**, mg **sawa**, lo **kōt̄ō**, lo **sowa**, te **kāt̄ā**, te **suwa**.

among *adv.* .

and *conj.* ay **ko**, cq **ko**, ja **ko**, do **ko**, hu **kau** $^{B.}$ so **ku**, ix **ku**, lo **ko**, te **ku**.

animal *s.* mz **ʧu**, ay **ʧu**, cq **ʧu**, ja **ʧu**, do **ʧu**, hu **ʧu** $^{B.}$ ji **ʧu**, so **ʧu** $^{B.}$ ix **ʧu**, mg **ʧu**, lo **ʧi**, te **ʧu** mz **ʧu**, ay **ʧu**, cq **ʧu**, ja **ʧu**, do **ʧu**, hu **ʧu** $^{B.}$ ji **ʧu**, so **ʧu** $^{B.}$ ix **ʧu**, mg **ʧu**, lo **ʧi**, te **ʧu**.

another *pron.* ay **ji?nku**, cq **ʃinku**, lo **ʃinku**.

ant *s.* mz **nī?ju**, ay **nī?iu**, cq **nti?iu**, ja **nī?iu** $^{L-M.}$ do **ndi?iu**, hu **nī?ju** $^{B-L.}$ ji **nī?ju**, so **nī?ju** $^{B-L.}$ ix **nti?iu**, mg **nī?ju**, lo **nī?ji**, te **nīju**.

apart *adj.* ay **ʃi**, cq **ʃi**, ja **ʃī** $^{M.}$ do **ʃi**, hu **wha?aʃī** $^{L-H-M.}$ so **ʃī** $^{H.}$ ix **ʃi**, lo **ʃi**, te **ʃi**.

appearance *s.* cq **tankōkū**, ja **watekū** $^{L-L-M.}$ do **watekū**, hu **ʧū** $^{L.}$ ji **ʧī**, so **ʧū** $^{LM.}$ lo **takokhi?inī**, te **kuatekū**.

1

Figure 4.3: A sample of an English-Mazatec *microdictionary*, according to Kirk's data (1966)

Mazateco

boca *s.* mz **tsu?a**, mz **ts?wa**, mz **ntsuwa**, ay **ntsu?wa**, cq **ntsuwa**, ja **nju?wa?3–1)**, ja **ju?wa** $^{L-H.}$ do **nju?wa**, do **ju?wa**, hu **nts?ua**, hu **ts?ua**, ji **ntsu?wa**, ji **ts?a**, so **ts?ua**, ix **nts?ua** ˙ ix **ntsu?wa?a** ˙ mg **ntsuw?a**, mg **ts?a**, lo **ntsu?wo**, te **ntsuwa**, te **tsuwa**.

bonito (de un sonido) *adj.* ay **ntathī**, ay **ntatsu**, ja **tshothī** $^{M-H.}$ ja **tshotsu** $^{M-M.}$ do **tshothī**, do **tshotsu**, hu **nāská thí** , hu **nāskatsu** $^{L-H-M.}$ ix **tshatsu**, te **tsuthī**, te **ntatsu**.

bordado *s.* ja **ʃantusinē** $^{M-M-M-M.}$ hu **?intu** $^{L-BL.}$ ji **?intu**, mg **?intu**, lo **ʃojintw**, te **ʃantu**.

borracho *adj.* mz **ʧ?i**, ay **ʧ?i**, cq **ʧi?iʧa**, ja **ʧ?i**, do **ʧ?i**, hu **ʧ?i** $^{B.}$ ji **ʧ?i**, so **ʧ?i** $^{B.}$ ix **ʧ?i**, mg **ʧ?i**, lo **ʧi?i**, te **ʧi**.

bosque *s.* mz **hŋā**, ay **nkihŋā**, cq **nkihŋāʧa**, ja **hŋāʧa** $^{H-H.}$ do **hŋāʧā**, hu **hŋā** $^{H.}$ ji **hŋā**, so **hŋā** $^{MH.}$ ix **nk?ihŋāsʧa**, lo **kihŋōʧo**, te **ihŋāsʧa**.

bravo *adj.* hu **tūā** $^{MH.}$.

brazo *s.* mz **thja**, ay **thia–**, cq **thia–**, ja **thia** $^{M.}$ do **thia**, hu **tsua**, ji **tsha**, so **thja**, ix **thianā**, mg **thia**, lo **tsho–**, te **tsha–**.

brillante *adj.* ay **?nā**, cq **tiwate**, ja **tiwatei** $^{H-M-M.}$ ja **?nā**, do **tiwatei**, do **?nā**, hu **whate**, hu **ʧhū?na:ā** $^{BM-B.}$ so **?nā** $^{B.}$ ix **kihwate**, ix **?nā**, mg **hwate**, lo **hwota**, te **tiwute**.

brillar *v. imf.* ay **wate**.

broda *v. 3s.* mz **tikhuja**, ay **tikhoja**, cq **tihwuja**, ja **tikhuja** $^{H-M-M.}$ do **tikhuja**, hu **khūja**, so **hūha** $^{L-ML.}$ te **tikuɲā**.

brotando *adj.* ay **ki?nti**, cq **ki?nti**, ja **khinti** $^{M-H.}$ do **khinti**, hu **?nti** $^{H.}$ ix **khinti**, mg **nt?i**.

brote *s.* ay **ʧ?u–**, ja **nāʃu?iu** $^{L-M-L.}$ do **nāʃu?iu**, hu **ʧ?u** $^{B.}$ ix **nāʃu?iu**, mg **ʧū**, lo **ʧi?a**, te **ju**.

brujería *s.* mz **the**, ay **the?e**, ja **th?e** $^{M.}$ do **th?e**, hu **thi?e**, so **the?e**, ix **the?e**, mg **th?e**, lo **tha?a**.

bueno *adj.* mz **ntathī**, ay **nta**, cq **nta**, ja **nda** $^{M.}$ do **nda**, hu **nta** $^{BL.}$ so **nta** $^{LH.}$ ix **nta**, lo **ntoho**, te **nta thī**.

buho *s.* mz **ʧohnū**, ay **thiuhŋū**, ay **ntekha**, cq **hunū**, cq **teikha**, ja **khunū** $^{H-H.}$ ja **tekha** $^{M-M.}$ do **ʧakhuhnū**, do **tekha**, hu **hnū** $^{H.}$ ji **hnū**, so **hnū** $^{MH.}$ ix **hnū**, mg **hnū**, lo **hnī**, te **hnū**.

busca *v. 3s.* ay **tihwinkankishai**, cq **kuankasei**, ja **tihwingishai** $^{H-M-M-M.}$ do **tihwingishai**, hu **washai** $^{MH-M.}$ ji **kihnʃishai**, so **tinʃitshai** $^{B-B-H.}$ ix **ts?ahwinkisai**, mg **hwaso**, lo **kowikotshe?e**, te **khuinʃishe**.

caballo *s.* mz **nā** , ay **nā** , ja **ndhē**, do **ndehe**, so **ntehe** $^{B-B.}$ ix **ntihe**.

cabaña *s.* mz **jashei**, ay **jashai**, cq **jasei**, ja **jashai** $^{H-M-}$ do **jashai**, hu **jashai** $^{H-L.}$ so **jatshai**, ix **jise**, ix **jase**, lo **jotshe**, te **?itshi**.

cabello *s.* mz **ntsha**, ay **ntsha**, cq **tsha**, ja **ntsha**, ja **tsha**, do **ntsha**, do **tsha**, hu **ntsha** $^{B.}$ ji **tsha**, so **tsha** $^{B.}$ ix **ntsha**, mg **tsha**, lo **tsho**, te **tsha**.

5

Figure 4.2: A sample of a Spanish-Mazatec *microdictionary*, according to Kirk's data

Mazateco

hīũ DO *adj.* dark; oscuro ← pMaz ***hŋũ**^{H}

hi?i CQ *s.* appearance of; aparencia ← pMaz ***khi?i**

hi?ntu LO *v. 3s.* rots; pudre ← pMaz ***?ntu**

hja MZ *s.* hawk; gavilán ← pMaz ***ʃha**^{B}

hje MZ *s.* sin; pecado ← pMaz ***ʃhe**^{H}

hje MZ *adj.* big; grande, alto ← pMaz ***ʃhe**

hjehje TE *p.p.* finished; acabado ← pMaz ***hehe**

hje? MZ *s.* corn; maíz ← pMaz ***ʃhe?e**$^{B-B}$

hjū^{M} HU *adj.* willing; intencionando ← pMaz ***hju**

hju HU *adj.* quiet; tranquilo ← pMaz ***ʃhiu**

hju JI *adj.* quiet; tranquilo ← pMaz ***ʃhiu**

hka MZ *adj.* blind; ciego ← pMaz ***tjka**^{B}

hka HU *s.* bald; pelón, calvo ← pMaz ***tka**

hki MZ *s.* medicine; medicina ← pMaz ***ntjki**^{HL}

hku^{B} HU *s.* head; cabeza ← pMaz ***tku**^{B}

hku JI *s.* head; cabeza ← pMaz ***tku**^{B}

hkue HU *adj.* rough; grueso ← pMaz ***tkwhe**

hkunāhmāntsha$^{B-B-H-L}$ HU *s.* finger tip; punta del dedo ← pMaz ***tkuntsha**$^{B-L}$**/*tkuntsa**$^{B-B}$

hkwa MZ *s.* snail; caracol ← pMaz ***tjkwa**^{H}

hmā MZ *adj.* black; negro ← pMaz ***hmā**^{M}

hmā AY *adj.* black; negro ← pMaz ***hmā**^{M}

hmā CQ *adj.* black; negro ← pMaz ***hmā**^{M}

hmā^{M} JA *adj.* black; negro ← pMaz ***hmā**^{M}

hmā DO *adj.* black; negro ← pMaz ***hmā**^{M}

hmā^{M} HU *adj.* black; negro ← pMaz ***hmā**^{M}

hmā^{H} SO *adj.* black; negro ← pMaz ***hmā**^{M}

hmā IX *adj.* black; negro ← pMaz ***hmā**^{M}

hmā MG *adj.* black; negro ← pMaz ***hmā**^{M}

hmā TE *adj.* black; negro ← pMaz ***hmā**^{M}

hmēnī$^{H-L}$ HU *pron.* what; ¿qué? ← pMaz ***–nī**

hmī^{B} AY *pref. pers.* pers prefix plur; prefijo pers. plur. ← pMaz ***hmī–**^{B}

hmī DO *pref. pers.* pers prefix plur; prefijo pers. plur. ← pMaz ***hmī–**^{B}

hmi– DO *s.* brother of spouse; cuñado ← pMaz ***ʧha**^{B}

hmīnti^{H-L} JA *adj.* sterile; estéril ← pMaz ***?nti**^{B}

hmīnʧhī$^{L-H}$ JA *s.* women, wives; mujeres ← pMaz ***nʧhī**^{H}

hmīnʧhī DO *s.* women, wives; mujeres ← pMaz ***nʧhī**^{H}

hmīnʧhīska^{L-H-H} JA *v. 3s.* plays; juega ← pMaz ***ska**^{H}

6

Fig. 4.4: A sample of a Mazatec *micro-dictionary*, according to *municipios*

The *Mazatec Geolinguistic Atlas* differs from the s.c. "linguistic atlases" currently available in Middle America, which do not provide geolinguistic data, but are actually language demography maps.

Figure 5.1 shows a standard grey coloured map, sorted on the basis of a Spanish lemma ("alto" = 'big, high'). In this case, lexical polymorphism with two simple adjectival roots *he* and *ʔnka/nkʔa/nkaʔa* versus composed adjectival stems such as *ntahe, tsanka, heikuʔnka* appears far more explicitly than in KIRK's original tables as shown previously in Figure 2.1.

The programme provides the user with 823 maps based on the English translations of the items, and 875 maps ordered according to the Spanish translations of KIRK's English glosses.

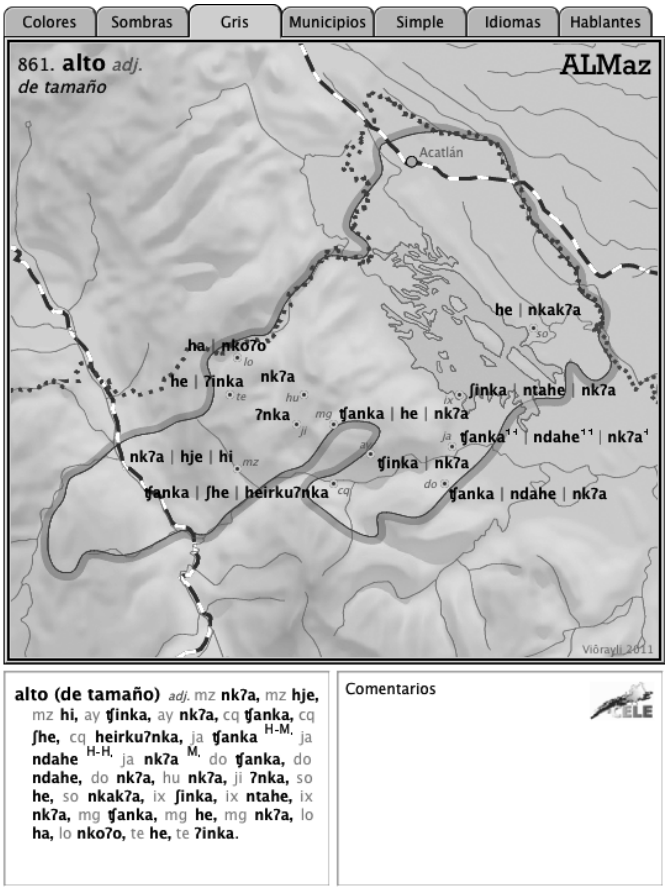

Figure 5.1: A grey style base map, with Spanish lemma for "big, high"

The 731 etyma proposed by KIRK are visualised on the corresponding maps. KIRK had an elaborate theory of Mazatec grammar and lexicon, mostly based on KENNETH PIKE's seminal work on the Huautla verb complex (PIKE 1948). Though map 216 (i.e. corresponding to KIRK's cognate set 216) deals with a nominal stem, KIRK had to elicit the root

kwja̱[6] with PMz **khwa* as a lemma by referring to derived concepts, such as "war", "gift", "law", "truth", etc. – he was faced with the same problem when eliciting compound verbs with a causative meaning, for which he had to ask for the forms for "builds up" as well as "purges" in order to get the root *tsi'ĩ corresponding to "make". The answers are actually easier to read in the ALMaz map Figure 5.2 than in Kirk's original work (cf. Kirk 1966: 295).

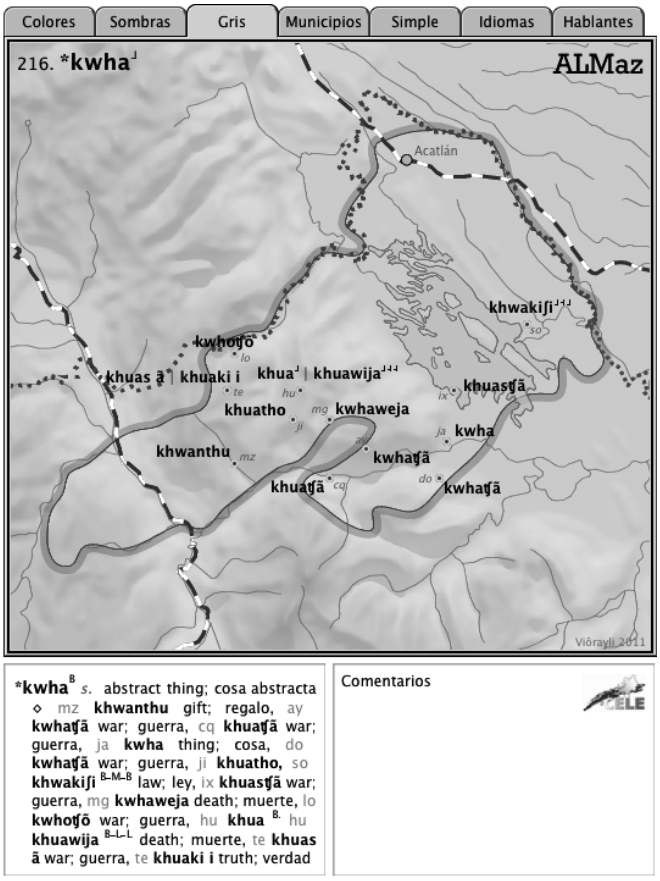

Figure 5.2: Simple and derived root for *kwha "abstract thing"

4.2. Generating typological traits

The rationalisation of the original data that was necessary to transform a simple text into a series of coherent and coordinated tables of a database, which can be easily enriched by tagging the entries with various bits of information in order to facilitate the search through the lexical items, or to study their phonetics, semantics and even their morphological properties in part, and eventually to analyse and classify them.

6 <j>, induced by Spanish "jota", stands for a breathy realization, and the underlined vowel has a low tone, according to current Mazatec script conventions.

Figure 6 shows the interface that allows the user to select phonological elements and their context in the roots (*etyma*): in this case "stressed A" in a palatal umlaut context, in an open syllable before N" – the example below is taken for a rhetoromance dialect, in this case.[7]

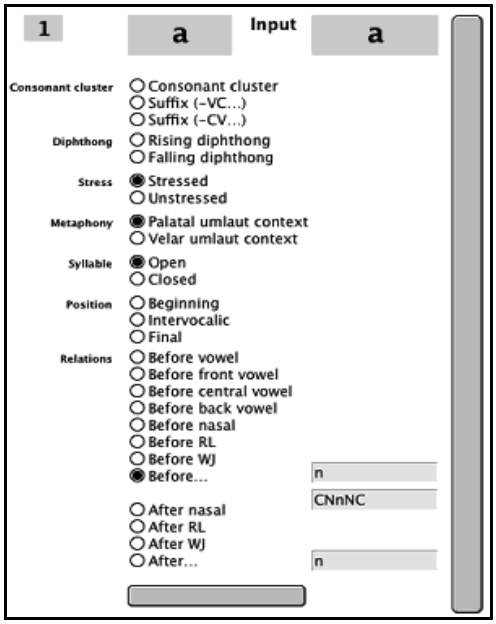

Figure 6: A query asking for local outcomes of roots starting by a nasal coronal sonorant /n/

Figure 7 shows the result of a query for the Ayautla reflexes of palatal nasal sonorants as onsets in Proto-Mazatec. The Target column at the right of the table points to the phonological changes, extracting the syllabic reflex in Ayautla for each cognate form.

Several traits pattern from this sample: denasalisation of the onset and palatalisation of the nucleus in numerals (*ɲãtu* > *jetu* = "seven", *ɲãtʃã* > *jetʃã* = "forty"), polymorphism of the nasal low vowel: *nũ-/nĩ-*, with vowel fronting *ũ* > *ĩ*, etc.

A beta version of a module of our database consists in an interface that shows the results of an algorithm that calculates and proposes possible phonetic evolution of single Proto-Mazatecan sounds (phonemes) for every spot on the map (see Figure 8).

This last cartographic example shows the interface hinting at the automatically identified outcomes for Protomazatec /ɲ/ as onset to a lexical root. The algorithm has taken into consideration all the items (local words) linked with the Protomazatec roots beginning by [ɲ] (between 10 and 17 words for each studied village), has proposed a regular outcome for each token ([ɲ], [n], [j], [i]) and has evaluated the accuracy of the answer (from 0.6 medium to 0.8 medium-high).

[7] This interface was initially developed on Dolomitan Ladin by VITTORIO DELL'AQUILA, before being applied to Mayan and Otomanguean languages in 2010.

Ayautla (San Bartolomé)

Id		pMaz	Forma	Target
1149	1	*ɲã*	tsaɲẽ	*tsaɲ-*
1150	1	*ɲãhã*	ɲãhã	*-ɲã-*
1151	1	*ɲãtu*	jetu	*-je-*
1152	1	*ɲãʔĩ*	nʔãĩtsu	*nʔã-*
1153	1	*ɲãʧãte*	jetʃãte	*-je-*
1155	1	*ɲũhu*	ɲũhũ	*-ɲũ-*
1156	1	*ɲũhũ*	ɲãhã	*-ɲã-*
1157	1	*ɲũmã*	nĩmã	*-nĩ-*
1158	0	*ɲũmẽ*	thiulimẽ	
1159	1	*ɲũnku*	nĩnkũ	*-nĩ-*
1160	1	*ɲũhmã*	nĩhmã	*-nĩ-*
1161	1	*ɲũhmẽ*	nĩhmẽ	*-nĩ-*
1163	0	*ɲũnku*	thiunĩnku	
1164	1	*ɲũnku*	nĩnku	*-nĩ-*

ListPhon_Auto

tsaɲẽ#ɲãhã#jetu #nʔãĩtsu#jetʃãte #ɲũhũ#ɲãhã#nĩ mã#nĩnkũ#nĩhm ã#nĩhmẽ#nĩnku# tsaɲ-#-ɲã-#- je-#nʔã-#-je-#- ɲũ-#-ɲã-#-nĩ- #-nĩ-#-nĩ-#-nĩ- #-nĩ-

Figure 7: Results of a query for the nasal palatal sonorant /ɲ/

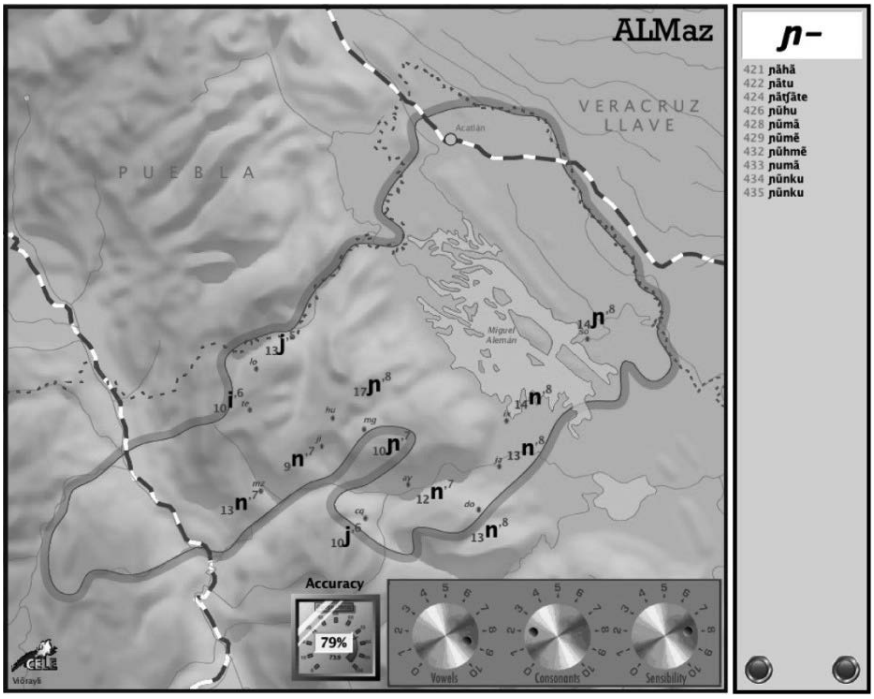

Figure 8

Not only the geolinguistic display of KIRK's data make linguistic categories and patterns more readable, but it also makes it possible to draw more isoglosses than GUDSCHINSKY was able to in her seminal survey of Mazatec dialect history (GUDSCHINSKY 1958b), and to suggest a finer-grained classification of Mazatec dialects.

Pondering 31 phonological variables, we obtained a more accurate classification of Mazatec dialects according to KIRK's data elicited in the sixties. The Mazatec area divides into three subareas according to climate, orography, agrarian patterns and cultural anthropology (BOEGE 1988): the Highlands (*Alta*), the Midlands (*Media*) and the Lowlands (*Baja*). These areas display three major agrarian types – Coffee in the Alta, Herd in the Media and Sugarcane in the Baja. This *Café/Ganado/Caña* complex, which has strongly interfered with the Meso-American Milpa system of Maize culture, had already started to prevail over the Milpa system when PAUL KIRK recorded his data. It has certainly had a very strong impact ever since on sociolinguistic networks, so that this classification, as earlier statements, is by no means definitive.

We could call cluster A "Northern Mazatec", cluster B "Southern Mazatec" and cluster C "Central Mazatec". This geolinguistic grouping is fairly congruent with GUDSCHINSKY's, but even more with the dialect intelligibility tests as reported by KIRK (1970) and CASAD (1965), especially in the case of group A, which is unexpectedly akin to the Baja dialect of Soyaltepec with the north-western dialects of San J. Tecoatl and San Lorenzo.[8] The Central Baja dialect (cluster B) matches the geographic concept of "Mazateca Media" (*Mid Mazateca*), and works as a transitional area between the Eastern Baja dialect and the Alta.

We shall mention a few criteria from our list of 31 phonological variables in order to sustain this classification. The following form a class of the more salient typological traits for Mazatec: syncope of homorganic vowels in *VhV* and *V'V* chains (the s.c. "rearticulated vowels"), pharyngalization of the high back vowel (*u* > *ɯ*), prenasalized stop denasalization (*nC* > *n*), lateralisation of a prenasalized palatal stop (*ntj* > *l*), neutralisation of breathiness (i.e. modalisation: *V̤* > *V*), retroflex affrication of oral palatal stop (*tj* > *tʂ*), aspiration of implosive sibilant (*sts* > *hts* = *tʂ̤*), preconsonantal assibilation of implosive prenasalized palatal stop (*ntjC* > *ʃC*), aspiration of implosive hushing sibilant (*ʃC* > *hC* = *Ç̤*). Among less typologically salient phenomena are processes of assimilation and dissimilation of vowel chains in disyllabic forms ($V_1CV_2 > V_1CV_1$ versus $V_1CV_1 > V_2CV_1$), diphthong reduction (*au* > *o*), mid-vowel lowering (*e* > *a*), nasal vowel fronting (*ũ* > *ĩ*), prothetic vowels, depalatalisation of palatal stops (tj/i > ti), affrication of breathy sibilants (*sh* or *s̤* > *tsh*, *tʂ̤*), nasal onset deletion (*n* > –), n. *Light* phenomena, such as vowel denasalization, voicing of prenasalized stops (*nt*, *nk* > *nd*, *ŋ*) and modalisation of creakiness, were given less weight in our ponderation of the 31 traits.

The classification of Mazatec dialects in the above complex of A, B and C clusters (Table 2) breaks the traditional bipartition between Highlands and Lowlands. The three sub-complexes broadly match major trends in the anthropological organisation of agrarian

[8] The San Lorenzo Cuaunecuiltitla dialect is associated with San J. Tecoatl as a derived form of the latter. Otherwise, what could easily be called *Cuaunecuiltiteco* is probably the most innovative Mazatec dialect as far as phonology is concerned. The link between both sub-groups (San J. Tecoatl & San Lorenzo on the one hand, and San Miguel Soyaltepec on the other hand) is twofold: by *mutual convergence*, as they surprisingly share a few strategic typological traits, and by *general divergence*, as they sum up a bulk of very idiosyncratic traits of their own, each in their domain (the concept of *shared structural excentricity* might apply here).

A	Alta north-West	San JeronimoTecoatl San Lorenzo
	North-East baja	Soyaltepec
B	South-West Cañada	Chiquihuitlán
	Central West Baja	Santo Domingo
		Jalapa
	Central East Baja	San Pedro Ixcatlán
		Ayautla
C	South-Alta	San Miguel (Huautla)
		Santa María Jiotes
	Central Alta	Huautla
	South-Western Alta	Mazatlán

Table 2: A classification of Mazatec dialects areas, according to 31 criteria

and cultural systems. It seems to follow a chain model of dialect grouping rather than the genealogical tree model (*Stammbaum*), as Jalapa and Santo Domingo in the Midlands whereas San Miguel and Santa María Jiotes in the Highlands form a cluster. We shall remain sceptical about the place of Mazatlan de Flores in this classification, which is biased by an empirical flaw in Kirk's data: it seems that in Mazatlán the author did not enjoy such good conditions for elicitation as elsewhere.[9] Our classification enhances the structural excentricity of several varieties, such as San Miguel Soyaltepec or San Lorenzo – in Kirk's survey of intelligibility (1970), Chiquihuitlán instead was considered as the most idiosyncratic dialect.

5. From a geolinguistic Atlas to a diasystemic approach of Mazatec typological traits

Table 3 highlights several typologically relevant processes, both from the standpoint of general phonology and mazatec dialectology, taking place in the Mazatec dialectical network. Mazatlán, Chiquihuitlán and Jalapa stand as diasystemic milestones in this matrix, where reflexes are qualified with labels in capital letters. The glosses read as follows: LAB = labial glide, PAL = palatalization (ex. u > i), RED = (vowel) reduction, MOD = modal voice and modalisation as neutralisation of breathiness or creakiness, OCCL = occlusion (e.g. #s > ts/_[+spread]), CONT = continuous (e.g. fricative), FLAT = flat, from sharp consonant (*tj* > *t*, i.e. depalatalization), SIMPLE = simplified CC cluster (e.g. sC > C), STRID = strident, VOC = vocalized.

[9] Intensive complementary fieldwork is planned for this important variety of Mazatec (with nearly 14000 speakers).

Item	**pMazatec**	English	**MZ Mazatlán**	**CQ Chiquihuitlán**	**JA Jalapa**
467	***su̯e** **LAB**	hot	tsi̯e OCCL, PAL, VOC	sue MOD, CONT	se̯ RED, CONT
465	***sa̯** **CONT**	bitter	tsa̯ OCCL	sa MOD	sa̯ CONT
577	***te̯e̯ (*t^he'e*)** **V̯'V**	magic, sorcery	te̯ V̯	– –	te̯ (*t^h'e*) V̯V̯
608	***t^jawa** **PAL**	white	tawa FLAT	tiawa PAL, VOC	tawa FLAT
610	***t^ja̯** **PAL, V̯**	arm	t^ji̯a STRID, VOC	t^ji̯a	ti̯a VOC
615	***t^ja̯u** **PAL, Au**	horn	–	t^ji̯o RED VOC, RED	
75	***tʃa̯u** **STRID, A̯u**	egg	tʃo̯ VOC, RED		

Table 3: A fragment of diasystemic analysis of the ALMaz data

Cognate set 467 (*shue = ***su̯e** > **MZ** *tsi̯e* **CQ** *sue* **JA** *se̯*) challenges Mazatec onset and nuclei distribution patterns as revisited in GOLSTON & KEHREIN (1998), and if the post-PIKE theory is correct, the labial glide should belong to the onset instead of forming a diphthong with e as head of the nucleus. Whatever may be the status of this labial glide, it may undergo palatalization (PAL) and therefore behaves like a nucleus instead of a consonant feature [+round], or it may vocalize (VOC). In CQ, breathiness is neutralized through modalization (MOD). CQ and JA instead maintain the continuity of ther sibilant. The highly marked breathy fricative *$s̤$ assumed for Proto-Mazatec may align with an unmarked affricate realization, such as *$t̤s̤$ in Mazatlán (***su̯e** > *tsi̯e*). Cognate set 465 (*shue = ***sa̯** > MZ tsa̯ CQ sa JA sa̯) confirms trends of the affricate alignment of a breathy sibilant in MZ, modalization of V̯ in CQ. Cognate set set 577 (**t^heʔe =** ***te̯e̯** > MZ te̯ JA t̤e̯) is a case study for general phonology, as it shows the intricate relationship between breathiness and creakiness in a s.c. rearticulated vovel (i.e. of the type VʔV). If KIRK's protoform **t^heʔe** is correct, a complex nucleus, bewilderingly rich in glottal autosegments, may simplify through breathy simplification (V̯) as in MZ, or make its inner complexity even more acute, as in JA, where we can see that syncope is not inhibited by glottal autosegmental conplexity, therefore licensing a breathiness-creakiness adjacency (t^hʔ) such as **t^heʔe =** ***te̯e̯** > t̤e̯ = *t^hʔe*.[10]

[10] For constraints on Otomanguean vowel complexity, see SILVERMAN (1997) and GOLSTON & KEHREIN (1998). Though such phonotactic complexes as breathiness + creakiness are observed in the World's languages and other Otomanguean languages, their robustness and the high level of gesture coordination implied in their realization make the issue fairly questionable.

Cognate sets 610–75 show how stridency (STRID) and breathiness interact (they are often correlates, inducing overcorrections in Mazatec script). Cognates 615 and 75 account for the reduction of what Pikean tagmemics would have considered a diphthong. However recent interpretations in terms of a redistribution of more or less autosegmental clues or features between onsets and nuclei (GOLSTON & KEHREIN 1998) would consider them as dependent from the onset. The low vowel component of the pseudo-diphthong *au, ao* would be simply a kind of diacritic for velarity. The reduction of the diphthong (*au, ao* > *o*) would therefore point at a finer typological process rather than trivial diphthong reduction: the velar autosegmental secondary feature of a palatal stop (615 ***t^{j}ạu**) or of a palatal affricate (75***tʃạu**) would therefore have undergone full vocalisation before it coalesced with the actual nucleus of the roots.

Any question one may wish to ask of the ALMaz database from a diachonic or from a typological standpoint can be answered via the tables and maps generated by the interfaces of the mapping architecture. Not only does the ALMaz provide maps, paradigms to any etymological and distributional queries, microdictionaries and glossaries, but it also gives clues for the falsification or the exploration of typological statements and hypotheses.

6. Conclusion

The ALMaz is more than a mere geolinguistic database compiled from a monograph: it is the matrix for a a linguistic atlas of Mazatec, and eventually, for other Otomanguean and Meso-american languages. It it part of a broader attempt to revisit data collected half a century ago, during the climax of empirical linguistics in Meso-America, when the structuralist or the functionalist methods (of which tagmemics is but a variant) still prevailed. The more one scrutinizes the output of this period, the more one is impressed by the quality and the amount of descriptive data which has been made available to modern linguists. Nevertheless, the more one dives into the data, the more one is convinced that the whole bulk of information is nothing more than a starting point. Fieldwork should be done again, many more spots on the map should appear and be systematically fed with considerably more elaborate and diverse data. The ALMaz will include 800 more cognate sets on inflectional morphology than in KIRK's Ph. D. dissertation and over 700 sentences in order to cover morphosyntax as well. But still, previous data is the matrix one starts with. Computerized geolinguistics can therefore be a powerful descriptive tool, a heuristic, critical, historiographical and epistemological method, as well as an outstanding opportunity to get back into the field.

References

BOEGE, ECKART (1988): *Los mazatecos ante la nación. Contradicciones de la identidad étnica en el México actual*. México, Fundo de Cultura Económica.

BULL, BRIAN (1984): Aspect formation of San Jeronimo Mazatec verb, in: *SIL Mexico Workpapers* 6, 93–117.

CASAD, EUGENE (1965): *Dialect intelligibility testing*. Dallas: Summer Institute of Linguistics.

COWAN, GEORGE (1965): *Some aspects of the lexical structure of a Mazatec historical text*. México D.F.: Summer Institute of Linguistics.

GOLSTON, CHRIS & KEHREIN, WOLFGANG (1998): Mazatec onsets and nuclei, in: *International Journal of American Linguistics* 64, 311–337.

GOLSTON, CHRIS & KEHREIN, WOLFGANG (2004): A prosodic theory of laryngeal contrasts, in: *Phonology* 21, 1–33. < http://zimmer.csufresno.edu/~chrisg/>

GUDSCHINSKY, SARAH (1958a): *Proto-Popotecan. A comparative study of Popolocan and Mixtecan*. (Indiana University Publications in Anthropology and Linguistics). Bloomington: Indiana University.

GUDSCHINSKY, SARAH (1958b): Mazatec dialect history, in: *Language* 34, 469–481.

JAMIESON, CAROLE (1982): Conflated subsystems marking person and aspect in Chiquihuitlán Mazatec verb, in: *International Journal of American Linguistics* 48(2), 139–167.

JAMIESON, CAROLE (1988): *Gramática mazateca del Municipio de Chuiquihuitlan, Oaxaca*. Mexico: Instituto Lingüístico de Verano, A.C.

JAMIESON, CAROLE (1996): *Diccionario mazateco de Chiquihuitlán*. Tucson: SIL.

KIRK, PAUL LIVINGSTON (1966): Proto-*Mazatec phonology*. PhD dissertation, University of Washington.

KIRK PAUL LIVINGSTON (1970): Intelligibility testing: The Mazatec study, in: *International Journal of American Linguistics* 36(3), 205–211.

KIRK, PAUL LIVINGSTON (1993): Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec, in: MATTINA, ANTHONY & MONTLER, TIMOTHY (eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson*. University of Montana, 235–250.

LADEFOGED, PETER & MADDIESON, IAN (1996): *The sounds of the world's languages.* Cornwall: Blackwell.

LÉONARD, JEAN LÉO (2010): Enquêtes exploratoires pour l'ALMaz (Atlas Lingüístico Mazateco). Elicitation croisée, entre typologie et codification d'une langue otomangue, in: *Géolinguistique* 11, 59–109.

LÉONARD, JEAN LÉO & KIHM, ALAIN (2010): Verb inflection in Chiquihuitlán Mazatec: a fragment and a PFM approach, in: MÜLLER, STEPHAN (ed.), *Proceedings of the HPSG10 Conference*. CSLI Publications (available on http://csli-publications.stanford.edu/).

MORENO, FROYLAN PÉREZ (2008): *Xujun én ntáxjo. Narraciones mazatecas con glosario Mazateco de Jalapa de Díaz y español*. Mexico: Instituto Lingüístico de Verano, A.C.

PIKE, EUNICE (1956): Tonally differentiated allomorphs in Soyaltepec Mazatec, in: *International Journal of American Linguistics* 22, 57–71.

PIKE, KENNETH & PIKE, EUNICE (1947): Immediate constituents of Mazatec Syllables, in: *International Journal of American Linguistics* 13, 78–91.

PIKE, KENNETH (1948): *Tone languages. A technique for determining the number and types of pitch contrasts in a language, with studies in tonemic substitution and fusion.* Ann Arbor, University of Michingan Press. NB: Rééditions 1949, 56, 57, 61.

SILVERMAN, DANIEL (1997): Laryngeal complexity in Otomanguean vowels, in: *Phonology* 14, 235–261.

SILVERMAN, DANIEL; BLANKENSHIP, BARBARA; KIRK, PAUL & LADEFOGED, PETER (1995): Phonetic structures in Jalapa Mazatec, in: *Anthropological Linguistics* 37, 70–88.

JEAN LÉO LÉONARD
Laboratoire de Phonétique et Phonologie
UMR 7018
CNRS/Sorbonne-Nouvelle
19, rue des Bernardins
75005 Paris
FRANCE
jeanleoleonard@yahoo.fr

VITTORIO DELL'AQUILA
Fjärilsvägen 9A5
65230 Vasa
FINLAND
viorayli@gmail.com

ANTONELLA GAILLARD-CORVAGLIA
37, bd. du Segrais, Apt. 28
77185 Lognes
FRANCE
antoc75@gmail.com